AMENDMENTS TO THE SPECIFICATION

[0015] (Currently Amended) Fig. 6 is a PCURVE™ ~~pCurve~~™ for a mixture dilution trends in accordance with the teachings of the present invention.

[0016] (Currently Amended) Fig. 7 is a flow chart illustrating an example of steps that may be taken to generate a PCURVE™ ~~pCurve~~™ such as shown in Fig. 6.

[0017] (Currently Amended) Fig. 8A shows an example of a T-CHART™ ~~T-chart~~ that may be used to identify significantly expressed genes using clone groups.

[0019] (Currently Amended) Fig. 8C, in comparison shows the same experimental data from Fig. 8B, having been plotted in a T-CHART™ ~~T-chart~~, according to the present invention, after taking noise factors into consideration. ~~consideration.~~

[0020] (Currently Amended) Fig. 9 is a flow chart illustrating steps that may be taken to distinguish differentially-expressed genes using the T-CHART™ ~~T-chart~~ of Fig. 8A in accordance with one embodiment of the present teachings.

[0026] (Currently Amended) A "PCURVE™ ~~pCurve~~™," as used herein, refers to a sorted p-value profile of a series of statistical, hypothesis-driven evaluations.

[0027] (Currently Amended) A "T-CHART™ ~~T-chart~~", as used herein refers to data re-plotted by coordinates, scaled in terms of noise units, so that statistical significance is more readily visually apparent.

[0059] (Currently Amended) As mentioned above, there may be more than 30,000 genes in a typical heterogeneous tissue sample and a scaled/corrected p-value for each gene can be calculated following the flow chart 500. A reliable p-value requires a sufficient population of samples taken from the heterogeneous tissue sample, where each sample may have its own mixture ratio of the two types of tissue. Another way of providing such population of samples can be mixing two types of tissue at controlled mixture ratios. For example, one can consider

a series of microarrays over changing condition, e.g., the Gene Logic mixture dilution series, where the hybrid solution goes incrementally from 100% liver tissue to 100% CNS (central nervous system) cell line. As genes can be expressed differently in the two types of tissue, a p-value for each gene expression profile and the trend profile can be calculated. Then, as disclosed in one embodiment of the present teachings, the p-values can be sorted and plotted in logarithmic scale to generate a curve, which may be referred to as a "PCURVE™ ~~pCurve™~~".

[0060] (Currently Amended) Fig. 6 is a plotted curve 600 (e.g., PCURVE™~~p-Curve~~) of sorted p-values against the ranks of the p-values based on the order of the sorted p-values from highly-significant, low p-values to larger, less-significant p-values. Each p-value, as statistically calculated, represents the probability that a response signature profile does not match a specified test signature profile defined by a template and/or clustering. However, a multiplicity of coincident p-values will stochastically produce some optimistic results. Hence, the smallest p-values forming the steep part of the PCURVE™~~pCurve~~ are the most reliable. In this example, curve (PCURVE™ ~~pCurve~~) 600 is for a Gene Logic mixture dilution series of liver tissue and CNS cell line. Curve 600 can be used to identify genes behaving differently between those two types of tissue. For example, the first 6,000 genes in Fig. 6 show a "very significant difference" (or, equivalently p-value ≤ 0.01), which may imply that the first 6,000 genes are related to the CNS cell line.

[0066] (Currently Amended) A "T-CHART™ ~~T-chart™~~" 800 (or, equivalently a scatter plot) of gene expression levels scaled by noise as obtained by replicates of measurements may be used to distinguish genes that have true differential expressions from those that might appear to be differentially expressed when plotting one value per gene, but which may not be truly differentially expressed when taking noise associated with the signal into consideration. Fig. 8A is a representation of a T-CHART™ ~~T-chart~~ 800 for gene expression levels of two types of tissue, type A and B, in a logarithmic scale. Typically, one of the two types of tissue may be a reference tissue, such as healthy tissue, while the other may be a diseased tissue. Each data point of the plot 800 corresponds to one replicate of measurement for a gene.

[0070] (Currently Amended) Figs. 8B-8C show a comparison between a plot 8000 (Fig. 8B) of gene expression levels from a red channel (LnRed) of a two-channel microarray platform plotted against gene expression levels for the same genes on a green channel (LnGreen) in a logarithmic scale. Chart 800' (Fig. 8C) shows a T-CHART™ ~~T-chart~~ of the same data, after noise-normalizing the data in the manner described above. The data points that are lighter in shade are those that were determined to be differentiated. Thus, in comparing these charts, it can be observed that some of the data points which might appear to show differentiated genes (e.g., 8012, 8014) are actually determined to not be significantly differentiated (e.g., 812, 814) when accounting for noise factors. In contrast, data point 8016 appears to be differentiated, and is also determined to be differentiated (816) after accounting for noise factors.

[0071] (Currently Amended) As mentioned, a T-CHART™ ~~T-chart~~ 800 is presented in a logarithmic scale. In a typical assay of biological study, the gene expression levels are generally plotted in logarithmic scale for both statistical and biological reasons. From a statistical standpoint, noise levels are usually approximately proportional to the signal level magnitudes. By taking the log of the readings, this homogenizes the noise levels relative to the signals, so that signal levels are not skewed by proportional log levels. From a biological viewpoint, the log of the signal is often proportional to the log of the stimulus, such as for example in the cases of vision, sound, and/or treatment versus response phenomena.

[0072] (Currently Amended) The a T-CHART™ ~~T-chart~~ 800 in Fig. 8A can be extended to high-dimensional space when gene expression levels are measured using multi-microarray apparatus. Based on the same reasoning applied to the analysis of gene expression in the a T-CHART™ ~~T-chart 800~~, a gene corresponding to a noise cloud in high-dimension space may be significantly expressed if the noise cloud does not overlap the high-dimensional diagonal.

[0074] (Currently Amended) At step 910, a T-CHART™ ~~T-chart~~ is generated, preferably in a logarithmic scale, using the measured and stored gene expression levels, in the manner described with regard to Fig. 8A above. Then, noise clouds generated from the plotting in step 910 are observed for each gene of interest, at step 912. Optionally, a forty-five degree diagonal line may be overlaid on the T-CHART™ ~~T-chart~~ 800 to aid in visibly determining whether any particular noise cloud is distinctly separated from the housekeeping genes (i.e.,

those genes substantially aligned with the forty-five degree diagonal which are considered to be neutral or not expressed). By observation or other analysis of the T-CHART™ ~~T-chart~~ 800, those genes corresponding to noise clouds that do not overlap with the diagonal of the T-CHART™ ~~T-chart~~ 800 are selected or identified as differentially-expressed genes in step 914. Optionally, the location of each point can be scaled by its particular noise factors to produce a chart of "standardized" points. The distance of each point from the diagonal becomes multiples of its particular noise factors. Hence, the distance automatically infers degree of overlap of noise with the diagonal, eliminating any need for plotting noise clouds.